ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Energy Procedia 37 (2013) 2011 – 2020

Energy Procedia

GHGT-11

Multivariate data analysis for parameters effect on CO_2 removal efficiency

Udara Sampath P.R.Arachchige[a,*], Neelakantha Aryal[a], Pramod Ghimire[a], Maths Halstensen[a], Morten Christian Melaaen[a,b]

[a]Telemark University College, Porsgrunn, 3901, Norway
[b]Tel-Tek, Porsgrunn, 3901, Norway

Abstract

In this paper, both the main effects and interaction effects of parameters on CO_2 removal efficiency were investigated. Flue gas stream data from a 500MW coal power plant has been used for the model development. The complete removal process is implemented in Aspen Plus with selected operating conditions and parameters using Monoethanolamine as solvent. The base case model is developed in Aspen Plus with specific parameter values to achieve 85% removal efficiency. The CO_2 removal efficiency variation with different parameters; such as number of stages, inlet solvent flow rate, lean loading, temperature of the flue gas and solvent stream, absorber packing height and diameter and absorber pressure are considered as the most important parameters for sensitivity analyses. The data collected from simulations were analysed using Principal Component Analysis, Principal Component Regression and Partial Least Square-regression. The correlation between variables were studied, which indicate that inlet solvent flow rate, absorber packing height and diameter, absorber pressure and temperature of the solvent stream are positively correlated with CO_2 removal efficiency whereas the lean loading and temperature of flue gas are negatively correlated with efficiency.

Keywords: Power generation; CO_2 emissions; Aspen Plus; Removal efficiency; Principal component analysis; Partial Least Square-regression

1. Introduction

Global warming and climate change effect, believed to be caused by the increased green house effect, has gained increasing attention in the last few years. Carbon released from large scale fossil fuel combustion

* Corresponding author. Tel.: +47-9428-41-16; fax: +47-355-75-001.
E-mail address: udara.s.p.arachchige@hit.no.

doi:10.1016/j.egypro.2013.06.081

is defined as the major emitting source today. Research studies on reducing green house gas emissions using CO_2 capture and sequestration has been implemented in recent years. The post combustion CO_2 capture via chemical absorption is still considered as a promising technology to achieve this goal. In order to make this process more economical, it is important to minimize the energy used in the regeneration section (re-boiler duty). The overall objective of this research study is to develop and implement a CO_2 removal model to find the most important parameters, and the corresponding effect on the removal efficiency.

The single parameter effect on removal efficiency has previously been studied [1]. Both the main effects (the effect of each individual parameter) and the interaction effects (interaction between two or several parameters) are discussed [2]. The objective in this study is to compare single parameter effects and multiple parameters effects on the CO_2 removal efficiency. The basic information related to the implemented model is described in the next section (Section 2).

The CO_2 removal base case model is developed for 500MW coal fired power plant flue gas. There is several parameters effect on CO_2 removal efficiency. The sensitivity analyses are performed to check the CO_2 removal efficiency variation with different parameters such as number of stages, inlet solvent flow rate, lean loading, temperature of the flue gas and solvent stream, absorber packing height and diameter and absorber pressure. By changing those parameters, CO_2 removal efficiency is calculated in Aspen Plus model. A total of 80 simulations are performed with different set of parameter values.

The data collected from simulation are analysed using Principal Component Analysis (PCA), Principal Component Regression (PCR) and Partial Least Square-regression (PLS-R). PCA can be defined as an orthogonal linear transformation that transforms the data to a new coordinate system. The transformation is defined according to the variance by any projection of the data and greatest variance is called first principal component, the second greatest variance is called the second coordinate, and so on [3]. PCR is considered as a powerful method for analysis of collinear data, which include both PCA and Multiple Linear Regression (MLR) [2].

2. Model Development

The flue gas stream data for 500 MW coal power plant is taken from Alie (2004), and implemented for removal process [4]. The composition of the flue gas is given in Table 1.

Table 1. Flue gas composition and parameters at inlet [4]

Parameter	Coal Fired
Flow rate [tones/hr]	2424
Temperature [°C]	40
Pressure [bar]	1.1
Major Component	Mol%
H_2O	8.18
N_2	72.86
CO_2	13.58
O_2	3.54
H_2S	0.05

The base case model is developed in Aspen Plus with specific parameter values to achieve 85% removal efficiency. Fig. 1 represents the flow diagram of the CO_2 removal process.

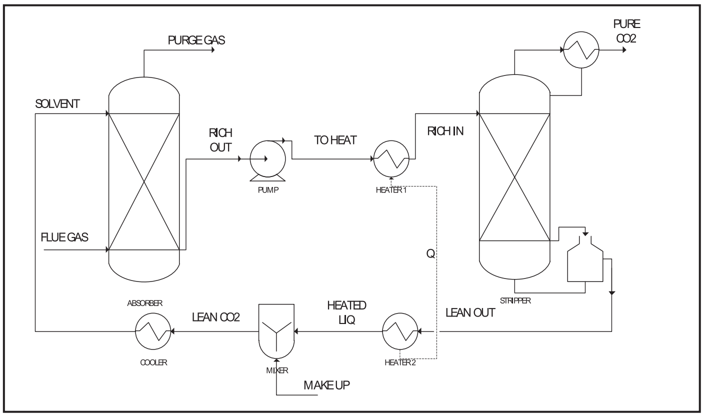

Fig. 1. Process Flow Diagram

The flue gas stream and solvent stream supply to the bottom and top of the column, respectively. Monoethanolamine (MEA) is used as the solvent for CO_2 capture. Chemical reactions take place in the packing material in the absorber column. A small portion of non-reacted CO_2 and other components leave from purge stream (PURGE-GA) deciding on overall capture efficiency. Rich out (RICH-OUT) is CO_2 abundant stream which is sent to other column for further processing. The main chemical reactions take place during the CO_2 removal process with MEA solvent [5].

The thermodynamic and kinetic data are selected according to the literature [6]. An open cycle complete removal process model is used and implemented to check the parameters' effect on CO_2 removal efficiency. The parameter values and operating range are tabulated in Table 2. A total of 80 samples were taken into consideration for parameter analyses.

The efficiency of the removal process is calculated using Equation 1.

$$\text{Efficiency} = \frac{\text{Total CO}_2 \text{ inflow} - \text{Purge gas CO}_2 \text{ outflow}}{\text{Total CO}_2 \text{ inflow}} * 100\% \qquad (1)$$

The interaction effects on the removal efficiency were found by varying several parameters according to a full factorial design scheme [2]. The data collected from simulation were analysed using PCA, PCR and PLS-R. The PCR and PLS models were validated using a so-called test set of independent data. A total of 80 samples were taken from the simulation out of which 30 were only used for validation of the PCR and PLS-R models. For the principal component analysis all 80 data samples were used. The commercial software The unscrambler were used for multivariate data analysis [7].

Table 1. Input parameter values in absorber column

Input parameter	Parameter condition (Fixed/Varied)	Base case value	Range of the parameter varied
Inlet flue gas (tones/ hour)	Fixed	2424.4	-
CO_2 content (mol %)	Fixed	13.58	-
Flue gas pressure (bar)	Fixed	1.1	-
Flue gas temperature-FT (ºC)	Varied	40	20-48
Packing material	Fixed	PALL type metal	-
Height of the packing-PH (m)	Varied	22	9-28
Diameter of the packing-PD (m)	Varied	16	8-20
Number of stages-NS	Varied	15	10-25
Inlet solvent flow rate-MF (tones/ hour)	Varied	7000	6000-16000
Solvent temperature-MT (ºC)	Varied	40	20-48
Solvent pressure (bar)	Fixed	1	-
Absorber pressure-AP (bar)	Varied	1	0.7-1.1
Solvent lean loading-LL %(mol CO_2/ mol MEA)	Varied	25	18-35
Solvent concentration (w/w)%	Fixed	25	-

After the calibration stage, the model must be validated based on independent data. Validation is needed in order to determine the model complexity in terms of number of principal components and also to get an estimate of the prediction performance of the multivariate model [2]. There are several validation techniques available such as cross validation, leverage correction validation and test set validation [2]. However, the test set validation method is used in this study.

3. Sensitivity Analysis

The temperature profiles in liquid and gas phase in the absorber column and the CO_2 loading profiles are analyzed for the base case model. The temperature bulge was seen at the top of the absorber. The magnitude of the temperature bulge is reached to 353K. Along the absorber, CO_2 loading is increasing and the maximum value reached 0.47 [mol CO_2/mol MEA] at the bottom. This section is devided into four different sub sections following matrix plot and scaling, PCA, PCR and PLS-R. It describes data pre-processing and explorative data analysis using PCA.

4.1 Matrix Plot

The histogram plot of the simulation data and the matrix plot can be used to check the necessity of scaling. Scaling is the method of weighting when parameters have different units and different variance. In order to ensure that all the data set roundly of same variance, pre-processing of the data before analysis is done. If any data set has higher variance, then the analysis might only explain the variation in the variable with higher magnitude. The important tool to decide whether the data set need scaling or not is called matrix plot (Fig. 2). According to the Fig. 2 the highest variance is given by MF (solvent flow rate). From the matrix plot, Fig. 2, it is clear that the data set has to be scaled and centred because of the variance of solvent flow rate is very high in comparison to the variance of other variables. Thus for making the loading plot to explain the variance of all the variables, the data set has to be scaled else the loading plot will explain mostly the variance in the solvent flow rate only.

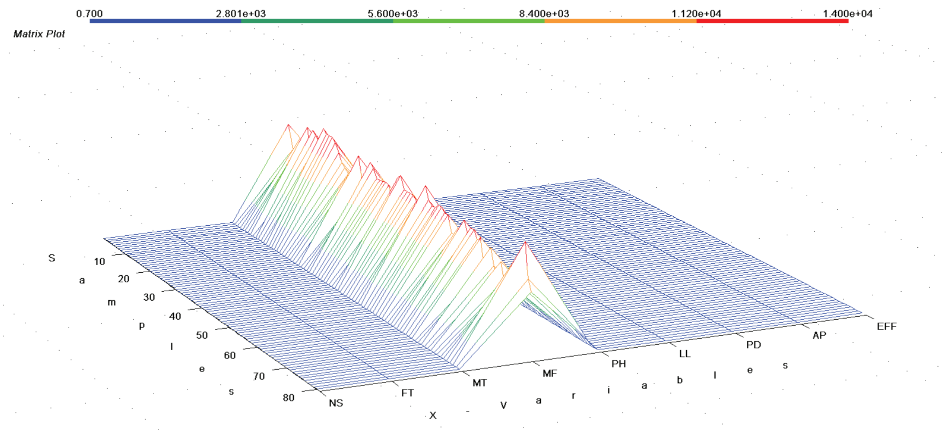

Fig. 2. Matrix plot of all the data samples and variables

4.2 Principal Component Analysis (PCA)

PCA is carried out to find the explorative model in order to observe the inter-dependability among the variables. The score plot is simply a relevant pair of score vectors plotted against each other. Score vectors are the coordinates of the objects projected down to the principal components.

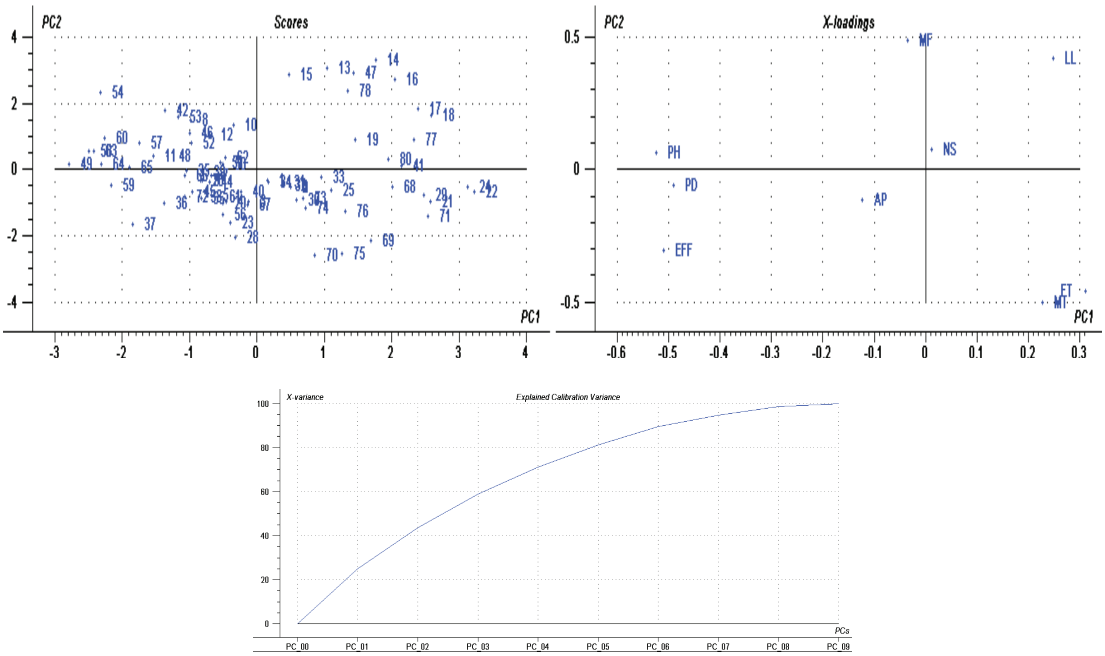

Fig. 3. Scores, Loading and Calibration Variance plot from PCA analysis without outliers: upper left figure shows score plot; upper right figure shows loading plot; bottom figure represents the calibration variance

Fig. 3 shows the scores and loading plot for PCA after deleting the outliers 43, 79 and 27. Outliers are selected which is given unexpected characteristics than rest of the data set. Score plot shows the location of the samples along each model component, and can be used to detect the sample pattern and understand the similarities of data. The samples along the Principal Components (PCs) PC1 and PC2 are considered because the PC1 and PC2 explain most of the structure and information in the raw data. PC1 explains 25% of the total data variance while 19% is explained by PC2. Sample number 22 has highest score along PC1 direction whereas sample 14 scored most in PC2 direction.

Analysing loading plot of Fig. 3, MEA Temperature (MT) and Fluegas Temperature (FT) and Lean loading (LL) are negatively correlated to the CO_2 removal efficiency variable. Packing height (PH), packing diameter (PD) and MEA flowrate (MF) are other variables which are positively correlated with the output variable i.e. efficiency (EFF). From explained calibration variance plot, it can be seen that 6 PCs are required to explain almost 90% variance in X variables and 8 PCs are required to explain 100% variance in X.

4.3 Principal Component Regression (PCR)

Multivariate calibration is also called multivariate modeling (X, Y). The Y vector contains the dependent variable whilst the X matrix is a set of independent variables. The multivariate model for (X, Y) is simply a regression relationship between X and Y established through multivariate calibration [2]. The model is used in the next stage for predicting new Y values. The matrix X and the corresponding Y are collectively known as calibration or training set. The training set is thus important to represent the future population. Mostly, two types of the calibration methods are used, namely PCR and PLS-R. After the model calibration, the model should be validated with the next data set of the same experimental setup, simulation data in this case. There are numerous mathematical validation techniques available like cross validation, leverage correction validation, test set validation, and however, the further analysis just focuses on test set validation. Test set validation is the only reliable one for validation.

The residual validation variance and predicted Y values are shown in Fig. 4. Test set validation method is used to validate the predicted model. The test set data (30 samples) are randomly selected. The slope of the predicted line is 0.60, the offset is 1.63 and the RMSEP (Root mean square error of prediction) is 8.64%. It is apparent from Fig. 4 that some included samples are extremed than the normal clustered values. So, the marked samples i.e. sample 27, 66, 72 and 78 have high score value regardless of others indicating that they don't match the whole clustered data, and are deleted considering outlier. After removing the outliers, the PCR is again performed. The score and loading along with the residual validation variance plot and the prediction plot are included in the Fig. 5.

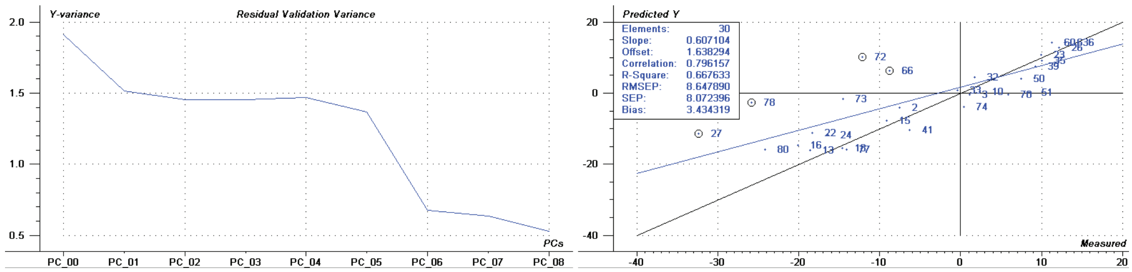

Fig. 4. PCR analysis of complete data set with efficiency as Y variable before removing the outliers; left side figure shows residual validation variance; right figure shows predicted Y.

From the residual validation variance plot, it can be seen that 6 PCs are required to explain the Y variance. Further, the slope of the predicted line is 0.79, the offset is 0.36 and the RMSEP (Root mean square error of prediction) is 4.57. This shows that there is improvement in these parameter analysis than before i.e with outliers. Again the X-Y loading plots can be seen in the Fig. 5 which shows that the MT, FT and LL are negatively correlated with the efficiency where as rest of the X variables are positively correlated to the efficiency. After all the refinements, the RMSEP is improved from 8.64% to 4.57%. Likewise significant improvement has been noticed in the slope of predicted line from 0.60 to 0.79 which can be seen from Fig. 4 and 5.

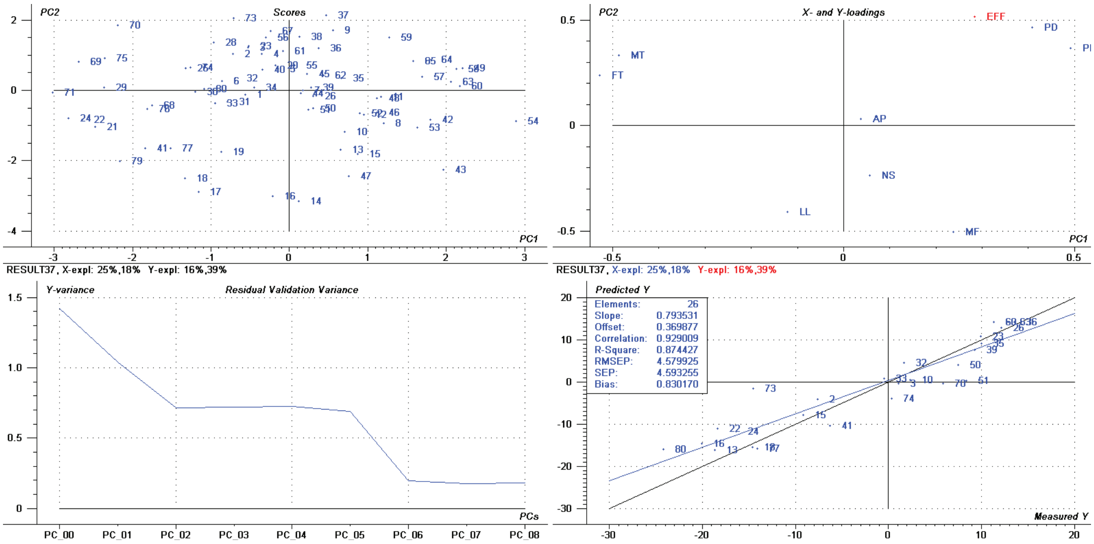

Fig. 5. PCR plots after removing all the outliers; upper left figure shows score plot; upper right figure shows loading plot; lower left figure shows residual validation variance; lower right figure represents predicted Y.

4.4 Partial Least Square-regression (PLS-R)

PLS-R uses the y-data structure, the y-variance, directly as a guiding hand in decomposing the X-matrix so that the outcome constitutes an optimal regression, precisely in the strict prediction validation sense [2]. In PLS, the components are not the principal components but the PLS components, however PC will be used for the simplicity [2]. The random test set taken for this PLS analysis are listed as: 2-3, 10, 13, 15-16, 18, 22-24, 26, 32-33, 35-36, 39,41, 50-51, 60, 63, 70, 73-74, 77, 80. Apart from these samples, other samples are used for calibration of model. Fig. 6 shows the PLS-R analysis for all the data sets before removing any outlier.

The PC1 explains 18% of X-variance whereas 70% of Y-variance. The number of PCs to explain optimal Y-variance seems to be 3. Slope of the predicted Y curve is 0.82 with offset 1.42% giving the RMSEP as 5.83%. The bottom figure shows weights of the regression coefficients which indicate that negative and positive impacts of parameters on CO_2 removal efficiency. According to that, LL and NS are negatively correlated with the removal efficiency and the rest of the other variables are positively correlated.

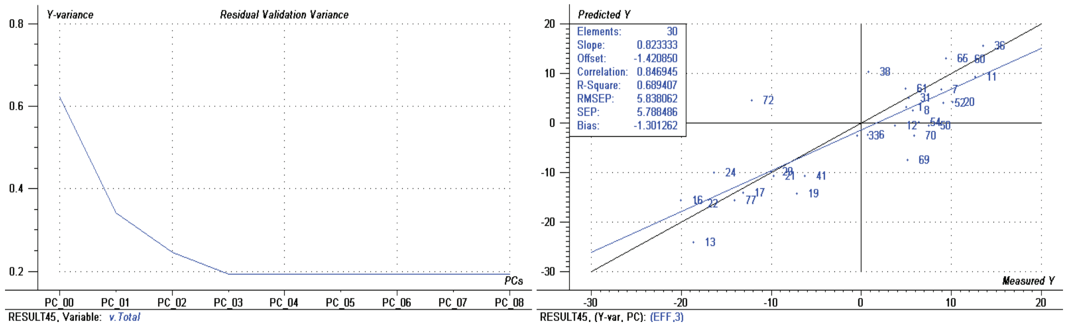

Fig. 6. PLS analysis of data set with 30 random test samples; left figure shows residual validation variance; right figure shows predicted Y.

The PLS-R is again performed without the marked outlier and the response obtained is analysed. Sample numbers 38 and 72 are identified as outlier and removed. Once more PLS-R is carried out and the obtained response is included in the Fig. 7. From the figure, it is seen that the slope has increased than before (Fig. 6) and there is considerable decrease in the RMSEP value which is the positive aspect as slope is expected to be 1 and RMSEP is expected to be 0 for making the perfect model. Further, the number of PCs required to explain sufficient Y variance is decreased to 2. The outliers in case of the PCR and PLS-R are different which is because the test set for the validation of data is taken in random from the total data set.

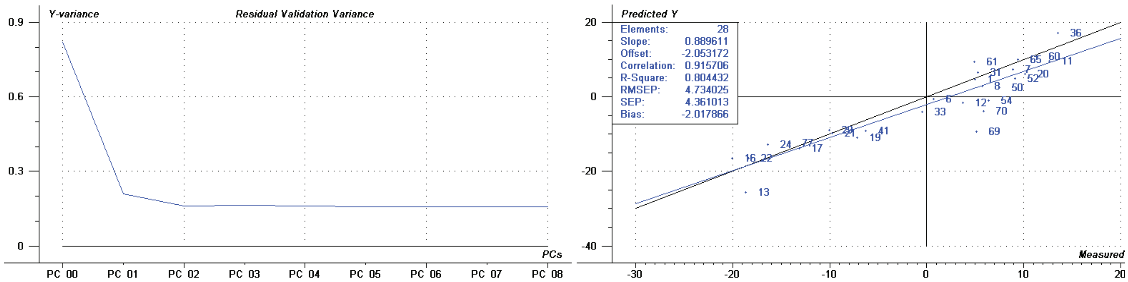

Fig. **7.** PLS model after deleting an outlier; left figure shows residual validation variance; right figure shows predicted Y.

From the regression coefficient chart (Fig. 8), it can be noted that the number of stages (NS) and Lean Loading (LL) have negative regression coefficients. NS and FT has no significant effect on the removal efficiency. LL has high negative correlation meaning, increase in LL with very small magnitude will cause a decrease in efficiency. In contrary to that, MF and PD have high positive correlation which means that the increase in MF and PD will cause an increase in the removal efficiency. From the regression coefficients plot, it can also be verified that LL, MF and PD are the most important X variables. MT, PH and AP have positive correlation with the efficiency and slightly less effect on efficiency than LL, MF and PD.

In PLS-R model, number of principal components (PCs) is less in comparison to PCR and thus it can also be noted that the PLS-R is faster and needs less number of PCs for explaining sufficient Y variance. Moreover, 18% of X variance and 69% of Y variance is explained only by PC1. X-loading weight in

Fig. 8 shows the importance of the variables in PC1 (blue), PC2 (red). For instance, NS, MF, PH and LL are the important variables for PC2. However, as almost 70% of Y variance is explained by PC1, our discussion and conclusion will be focused on the variables effect on PC1. For PC1 analysis, PH, LL and PD are the most important variables.

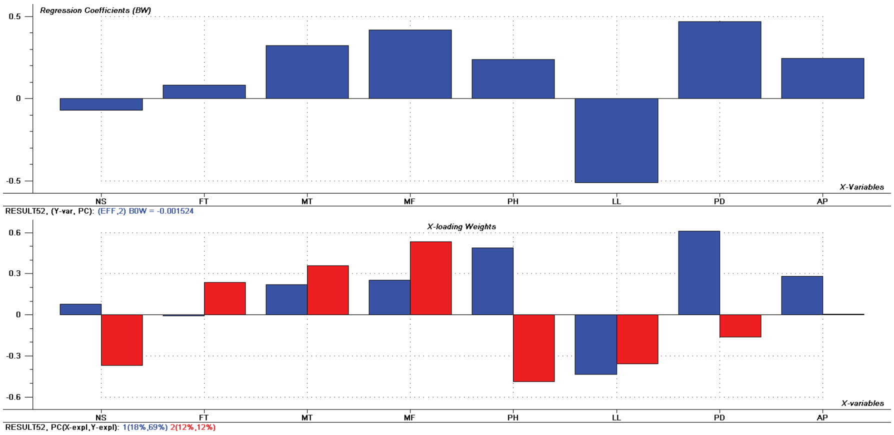

Fig. **8.** Regression Coefficient and X-loading weights plot; upper figure shows regression coefficients and bottom figure shows importance of the variables in PC1 and PC2.

5. Conclusion

The post combustion CO_2 capture model was developed and implemented in Aspen Plus. Correlations between the variables were studied, which indicate that inlet solvent flow rate (MF), absorber packing height (PH) and diameter (PD), absorber pressure (AP) and temperature of the solvent stream (MT) are positively correlated with the efficiency whereas the lean loading (LL) is negatively correlated with efficiency. From the regression coefficient plot in the PLS-R analysis, it can be noticed that inlet solvent flow rate, lean loading and packing diameter are the most important variables for removal efficiency. Number of stages and flue gas temperature are found to be less significant for the removal efficiency. Multivariate data analysis of the absorber column of the CO_2 capture plant is a promising technique for selection of optimal parameters to modify in order to achieve higher CO_2 removal efficiency. The single variable effect on efficiency was previously been studied with keeping other variables constant. Both the main effects (the effect of each individual parameter) and the interaction effects have been studied in this paper. Effect of the parameters on CO_2 removal efficiency is given the same conclusion in both cases. Single variable effect (by keeping other variables as constant value and change only one at once) is not good enough to understand the effect on that variable on removal efficiency. The regression coefficients can be used to develop the model that can predict the future variations with parameters.

References

[1] Arachchige USPR, Aryal N, Melaaen MC. Case study for flue gas separation of a coal fired power plant and parameters' effect on removal efficiency, Proceedings, APCRE'11 chemical engineering symposium, Beijing, China; 2011.

[2] Esbensen KH. *Multivariate Data Analysis – in practice.* 5.th Ed. (Oct. 2001), CAMO AS Publ. ISBN 82-993330-2-4; 2001.

[3] Martens H, Næs T. *Multivariate Calibration.* John Wiley & Sons Ltd., UK; 1989 (reprint 1994).

[4] Alie CF. CO_2 Capture with MEA: Integrating the Absorption Process and Steam Cycle of an Existing Coal-Fired Power Plant. Master Thesis, University of Waterloo, Canada; 2004.

[5] Michael AD. A model of vapour-liquid equilibria for acid gas-alkanolamine-water systems. PhD Thesis, University of Texas, USA; 1989.

[6] Freguia S. Modeling of CO_2 removal from Flue Gas with Mono-ethanolamine. Master Thesis, University of Texas, USA; 2002.

[7] Martens H, Karstang T, Næs T. *Improved selectivity in spectroscopy by multivariate calibration.* Journal of Chemometrics 1(4):201-219; 1987.